

Appl. No. 09/996,271
Amdt. dated February 1, 2005
Reply to Office Action of November 3, 2004

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

Claim 1 (Currently Amended): A system for processing an unstructured multimedia data file to provide information supporting user navigation of multimedia data file content, comprising:

a page layout analyzer to ~~determine~~ establish page layouts for each page of the unstructured multimedia data file, the analyzer ~~identifying~~ creating a structure ~~of~~ for each page layout including text sections and image sections of each page;

a content parser to identify previously unidentified text and image content of a data file, the content parser applying text extraction rules to identify text and identify a document structure, wherein context is defined for the identified text based on its associated document structure;

a black and white image processor for processing said identified image content to identify embedded text content by applying object templates, the image processor comprising a pixel smearing component reducing text to a rectangular block of pixels and an image filtering component for cleaning a smeared image;

a text sorter for parsing said identified text and said identified embedded text to locate text items in accordance with predetermined sorting rules;

a document processor for hyperlinking a plurality of keywords within the identified text and identified embedded text to a related document by creating an anchorable information unit file, wherein the plurality of hyperlinked keywords are anchorable information units; and

memory for storing a navigation file containing said text items and said anchorable information unit file.

Claim 2 (Original): The system of claim 1, wherein the navigation file links to at least one internal document object.

Claim 3 (Original): The system of claim 1, wherein the navigation file links to at least one external document object.

Claim 4 (Canceled).

Claim 5 (Canceled).

Claim 6 (Original): The system of claim 1, wherein the content parser applies pre-defined hierarchical rules for determining a level of identified text.

Appl. No. 09/996,271
Amdt. dated February 1, 2005
Reply to Office Action of November 3, 2004

Claim 7 (Canceled).

Claim 8 (Original): The system of claim 1, wherein the system refines a search resolution during a text identifying process to determine a location of the embedded text within an image.

Claim 9 (Original): The system of claim 1, wherein identified text comprises hyperlinks.

Claim 10 (Canceled).

Claim 11 (Canceled).

Claim 12 (Canceled).

Claim 13 (Canceled).

Claim 14 (Cancelled).

Claim 15 (Canceled).

Claim 16 (Canceled).

Claim 17 (Canceled).

Claim 18 (Canceled).

Claim 19 (Canceled).

Claim 20 (Canceled).

Claim 21 (Canceled).

Claim 22 (Canceled).

Claim 23 (Currently Amended): A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for creating an anchorable information unit file from an unstructured portable document format document, the method steps comprising:

analyzing a portable document format document to ~~determine~~ establish page layouts for each page of the portable document format document;
~~identifying~~ creating a structure ~~of~~ for each page layout including text portions and non-text portions for each page;

parsing the portable document format document into textual portions and non-text portions;
extracting structure from the textual portions and the non-text portions;
determining text within textual portions, and the non-text portions; and
hyperlinking a plurality of keywords within the textual portions and non-text portions to a related document by creating the anchorable information unit file, wherein the plurality of hyperlinked keywords are anchorable information units.

Claim 24 (Original): The program storage device of claim 23, wherein the step of parsing further comprises the step of differentiating color image content from black-and-white content.

Claim 25 (Original): The program storage device of claim 23, wherein the step of extracting further comprises the steps of:
determining a level for extracted textual portions;
associating the context with the text; and
pattern matching extracted text to the portable document format document to determine a context and a location.

Claim 26 (Original): The program storage device of claim 25, wherein the level is one of a paragraph, a heading and a subheading.

Claim 27 (Original): The program storage device of claim 25, wherein the step of pattern matching further comprises the steps of:
determining a median font size for the portable document format document;
comparing a font size of the extracted text to the median font size for the portable document format document and
determining a context according to font size.

Claim 28 (Canceled).